Article No. jmbi.1999.2945 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 291, 1091–1104

JMB

Sequence Homology and Structural Analysis of the Clostridial Neurotoxins

D. Borden Lacy[1] and Raymond C. Stevens[2]*

[1]*Department of Chemistry University of California at Berkeley, Berkeley, CA 94720, USA*

[2]*Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA*

**Corresponding author*

The clostridial neurotoxins (CNTs), comprised of tetanus neurotoxin (TeNT) and the seven serotypes of botulinum neurotoxin (BoNT A-G), specifically bind to neuronal cells and disrupt neurotransmitter release by cleaving proteins involved in synaptic vesicle membrane fusion. In this study, multiple CNT sequences were analyzed within the context of the 1277 residue BoNT/A crystal structure to gain insight into the events of binding, pore formation, translocation, and catalysis that are required for toxicity. A comparison of the TeNT-binding domain structure to that of BoNT/A reveals striking differences in their surface properties. Further, the solvent accessibility of a key tryptophan in the C terminus of the BoNT/A-binding domain refines the location of the ganglioside-binding site. Data collected from a single frozen crystal of BoNT/A are included in this study, revealing slight differences in the binding domain orientation as well as density for a previously unobserved translocation domain loop. This loop and the conservation of charged residues with structural proximity to putative pore-forming sequences lend insight into the CNT mechanism of pore formation and translocation. The sequence analysis of the catalytic domain revealed an area near the active-site likely to account for specificity differences between the CNTs. It revealed also a tertiary structure, highly conserved in primary sequence, which seems critical to catalysis but is 30 Å from the active-site zinc ion. This observation, along with an analysis of the 54 residue "belt" from the translocation domain are discussed with respect to the mechanism of catalysis.

Keywords: clostridial neurotoxin; botulinum neurotoxin; tetanus neurotoxin; translocation; X-ray crystallography

Introduction

The *Clostridium botulinum* organism produces seven immunologically distinct forms of botulinum neurotoxin (BoNT), designated A-G, from different strains (Simpson, 1989). BoNT/E and BoNT/F have been observed from *Clostridium butyricum* and *Clostridium baratti*, respectively (Aureli *et al.*, 1986; Hall *et al.*, 1985). Each toxin is synthesized as an inactive ~150 kDa single-chain protein. The protein is post-translationally proteolyzed to form the active dichain molecule in which the two chains, ~50 and ~100 kDa, remain linked by a disulfide bond (Figure 1(a)). The active dichain molecule is comprised of three ~50 kDa functional domains: binding, translocation, and catalytic (Montecucco & Schiavo, 1995). The binding domain comprises the C-terminal half of the ~100 kDa chain, while the translocation domain is located in its N-terminal half. The catalytic domain, a zinc-endopeptidase, is confined to the N-terminal 50 kDa chain.

These domains correlate with a three-step model for toxicity (Simpson, 1980). In the first step, the binding domain mediates interaction between the toxin and the presynaptic nerve terminal membrane (Dolly *et al.*, 1984). This interaction is thought to occur through both a ganglioside and protein receptor (Montecucco, 1986). Following

Abbreviations used: CNT, clostridial neurotoxin; TeNT, tetanus neurotoxin; BoNT, botulinum neurotoxin; VAMP, vesicle-associated membrane protein; SNAP, synaptosomal-associated protein.

E-mail address of the corresponding author: stevens@scripps.edu

0022-2836/99/351091–14 $30.00/0

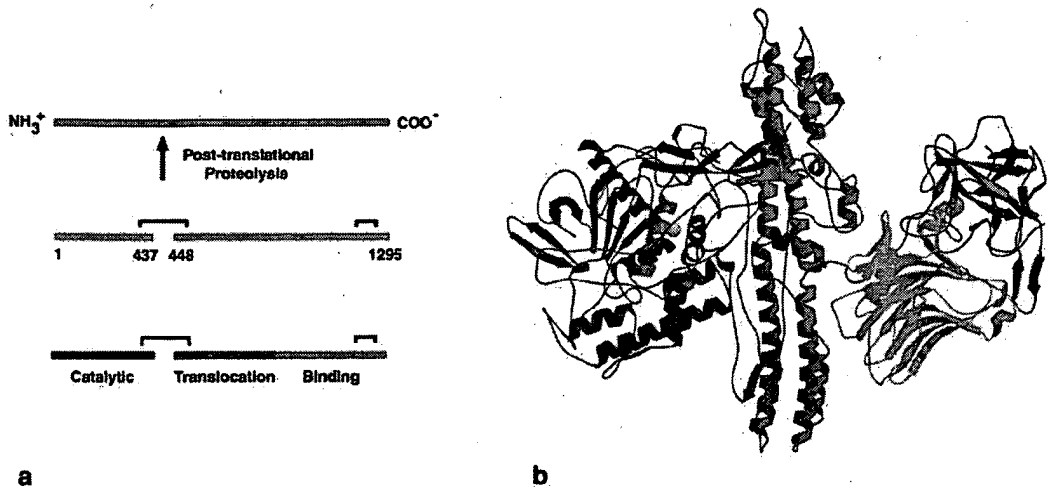

Figure 1. (a) The 150 kDa toxin is post-translationally proteolyzed to form the activated dichain molecule (the arrow in the top part of the Figure illustrates where proteolysis occurs). Two disulfides exist, one that links the two chains and the other in the C-terminal half of the binding domain. The three functional domains are each ~50 kDa and correspond to the catalytic domain (1-437), the translocation domain (448-872), and the binding domain (873-1295), where the numbers refer to the BoNT/A sequence. (b) A backbone trace of the BoNT/A structure with the catalytic domain in purple, the translocation domain in green, the N-terminal binding subdomain in pink, and the C-terminal binding subdomain in blue. The catalytic zinc ion is represented as a sphere with the helix containing the HEXXH zinc-binding motif in red. Helix α22 of the C-terminal binding subdomain is colored orange to show the relative orientation of the putative ganglioside-binding site to the rest of the molecule. The structure represents residues 1-431 and 450-1295, with residues 432-437 and residues 448-449 at the site of proteolytic cleavage being disordered. This Figure was generated using MOLSCRIPT and Raster3D (Kraulis, 1991; Merritt & Bacon, 1997).

binding, the protein is internalized by receptor-mediated endocytosis (Black & Dolly, 1986a,b). The second step is triggered by the lower pH of the endosome. Acidic pH is thought to cause a structural change in the translocation domain, allowing it to form a pore in the membrane. This pore allows for the translocation of the catalytic domain across this membrane, gaining access to its cytosolic target. The third step involves the cleavage of one of three proteins involved in synaptic vesicle membrane fusion. BoNT/B, BoNT/D, BoNT/F, and BoNT/G cleave distinct sites within a vesicle-associated membrane protein (VAMP, also referred to as synaptobrevin) (Schiavo *et al.*, 1992, 1993a,c, 1994; Yamasaki *et al.*, 1994a,b). BoNT/A and BoNT/E recognize and cleave distinct sites near the C terminus of SNAP-25 (synaptosomal-associated protein of 25 kDa) (Binz *et al.*, 1994; Blasi *et al.*, 1993a; Schiavo *et al.*, 1993a,b), while BoNT/C1 cleaves syntaxin and SNAP-25 (Blasi *et al.*, 1993b; Foran *et al.*, 1996). VAMP, SNAP-25, and syntaxin are collectively termed the SNARE proteins and have been shown to interact in a four-helix coiled-coil in a step thought to precede synaptic vesicle membrane fusion (Sutton *et al.*, 1998). Cleavage of any one of these three proteins prior to the coiled-coil formation disrupts neurotransmitter release at the neuro-muscular junction. The disruption leads to the flaccid paralysis observed in the disease botulism.

TeNT is produced by *Clostridium tetani* but shares ~65% sequence homology and ~35% identity with the BoNT serotypes. Tetanus toxin is a 150 kDa dichain molecule, which is thought to undergo a three-step mechanism of binding, translocation, and catalysis. The key difference is that while it binds at the presynaptic nerve terminal ending, it undergoes retrograde transport up the axon to act in the central nervous system. The molecular determinant of this alternate localization is unknown but is thought to lie in differences in the protein receptor. TeNT cleaves a site in VAMP, identical with that cleaved by BoNT/B (Schiavo *et al.*, 1992). However, given the different localization, the effect of this cleavage is to inhibit the release of inhibitory neurotransmitter and causes the spastic paralysis seen in tetanus poisoning.

The structure of BoNT/A was recently solved to 3.3 Å resolution using data collected from multiple crystals at 4°C (Lacy *et al.*, 1998). The structure (Figure 1(b)) showed that the binding domain was structurally similar to the TeNT binding domain (Umland *et al.*, 1997), and could be divided into two subdomains, an N-terminal β-barrel and a C-terminal β-trefoil fold. The translocation domain fold is markedly different from the folds observed in other toxins that undergo pore formation and translocation (Lacy & Stevens, 1998). Most notable are a kinked pair of α-helices, 105 Å in length, and a 54 residue "belt" that wraps around the

perimeter of the catalytic domain. The translocation domain occludes access to a large, negatively charged cleft leading into the active-site zinc ion of the catalytic domain. The zinc ion appears to be directly coordinated by His222, His226, Glu261 and by a water-mediated coordination through Glu223.

The CNT sequence divergence is substantial when one considers that the family of proteins has maintained striking similarities in a complex series of functions. In spite of the divergence though, it is reasonable to presume that the BoNT/A structure will be representative for all of the CNTs. Firstly, there is a greater conservation in predicted secondary structure than in the primary structure (Lebeda & Olson, 1994). Secondly, the structure of the binding domain overlays with the tetanus neurotoxin-binding domain with a root-mean-square deviation (rmsd) of 1.5 Å for 363 Cα atoms. Lastly, the functional requirements, especially those of acid-triggered pore formation and translocation, are rigorous and are not likely to tolerate significant structural changes. It is tempting to think that viewing the primary structure differences within the tertiary structure will lend insight into the molecular determinants of differences between the CNTs. The key differences seem to be an ability to discriminate between protein receptor molecules and in the enzyme specificity for one of eight cleavage sites in the three SNARE proteins. An analysis of the sequence conservation can also help identify regions required in the common functions of ganglioside-binding, pore formation, translocation, and catalysis.

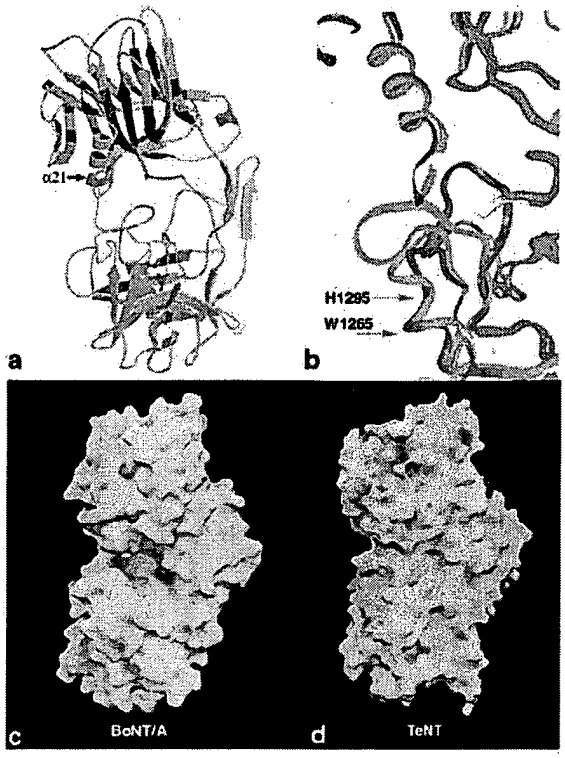

Figure 2. (a) The conserved residues of the BoNT/A-binding domain are colored purple, indicating that the N-terminal sub-domain (top) is more highly conserved than the C-terminal sub-domain (bottom). The second disulfide in BoNT/A, green, is located behind α22. This helix, with residues W1265 and Q1269 colored red, is thought to form part of the ganglioside-binding site. (b) A superposition of the BoNT/A (yellow) and TeNT (red) binding domains at the ganglioside-binding site and a loop difference between the two structures. (c) The two subdomains of the BoNT/A-binding domain are connected by α21 and form an electrostatically positive cleft in their interface. (d) The cleft in TeNT is relatively shallow and neutral, largely due to the loop difference shown in (b).

Results and Discussion

The binding domain

The structure of the TeNT-binding domain has been solved by independent investigators to 2.7 Å (Umland *et al.*, 1997; PDB code 1af9) and 1.5 Å (PDB accession code 1A8D). The main chains of the structures differ only in the orientation of two surface loops, located in crystal packing interfaces. As both structures were obtained in the same space group, these packing interface differences, along with the temperature and X-ray source differences are likely to account for the differences in resolution. The structure of BoNT/A at 4 °C, and now −170 °C, reveals the orientation of the binding domain and the accessibility of the surface loops in the presence of the translocation domain (in comparison to the TeNT-binding domain structure). The binding domain tilts away from the plane of the catalytic and translocation domains by ~60°. It projects away from the long axis of the translocation domain by ~40° in the structure from 4 °C data and by ~45° in the structure refined into frozen data. This observation is not surprising, given the relatively small interface between the binding and translocation domains. (The binding domain buries ~400 Å^2 of the translocation domain, while the translocation domain buries ~480 Å^2 of the binding domain.) The relative orientation of this domain could vary even more substantially under physiological conditions. The two sub-domains of the binding domain are linked by an α-helix (α21) and create a cleft in their interface (Figure 2(a) and (c)). The C-terminal subdomain buries ~500 Å^2 of the N-terminal subdomain, while the N-terminal subdomain buries ~540 Å^2 of the C-terminal domain. The contacts between the two subdomains are made through loops such that this interface could also be flexible, although the subdomain β-strands align almost identically in the three structures.

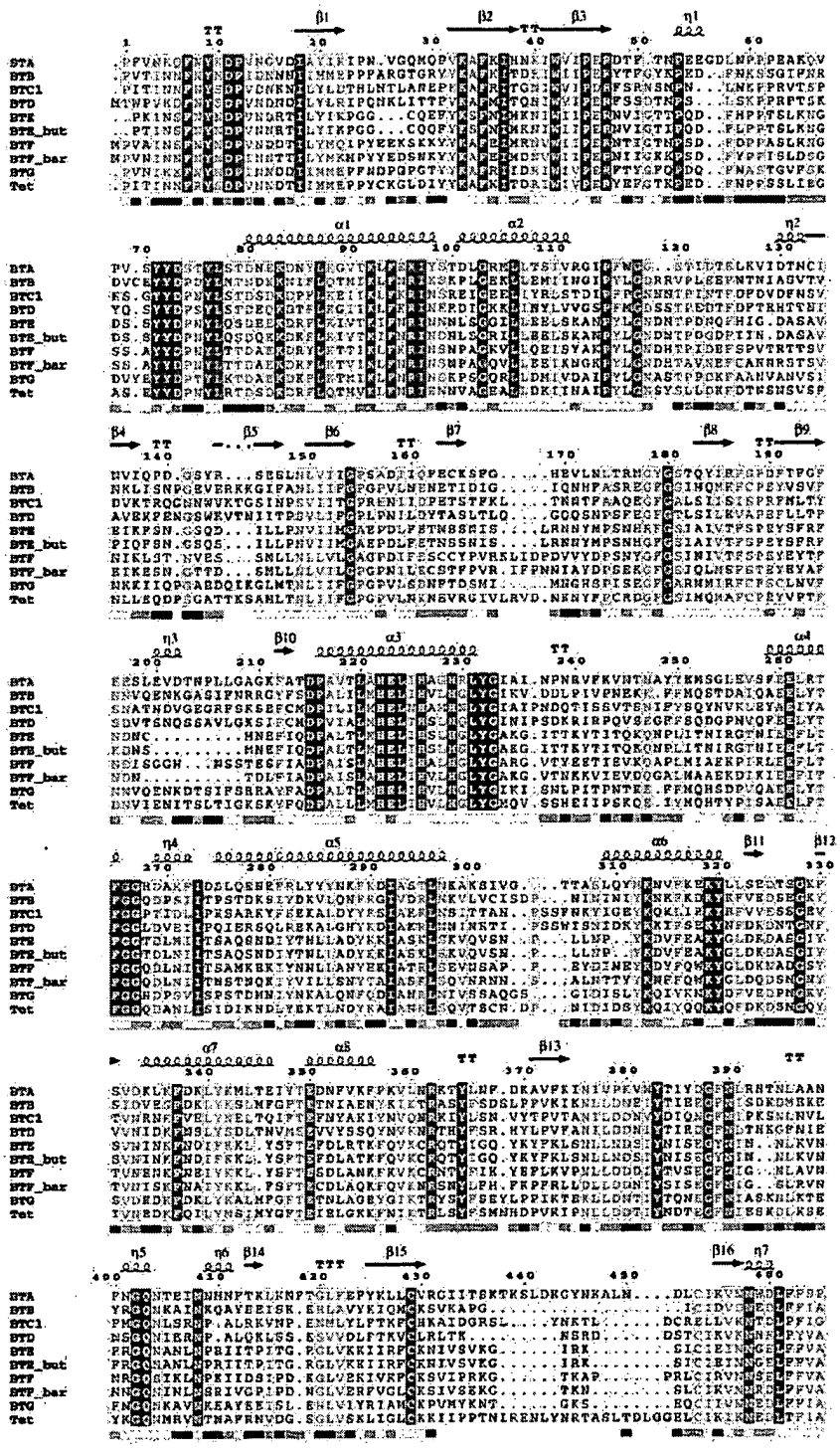

Figure 3 (*legend shown on page 1096*)

The alignment of BoNT/A with TeNT and the six other BoNT serotypes from the three identified organisms is shown along with the secondary structure assignments and solvent accessibility of residues in the BoNT/A structure in Figure 3. There is significantly greater sequence homology within the N-terminal binding subdomain as compared to the C-terminal subdomain (Figures 2(a) and 3). The majority of these highly conserved residues in the N-terminal subdomain, however, point

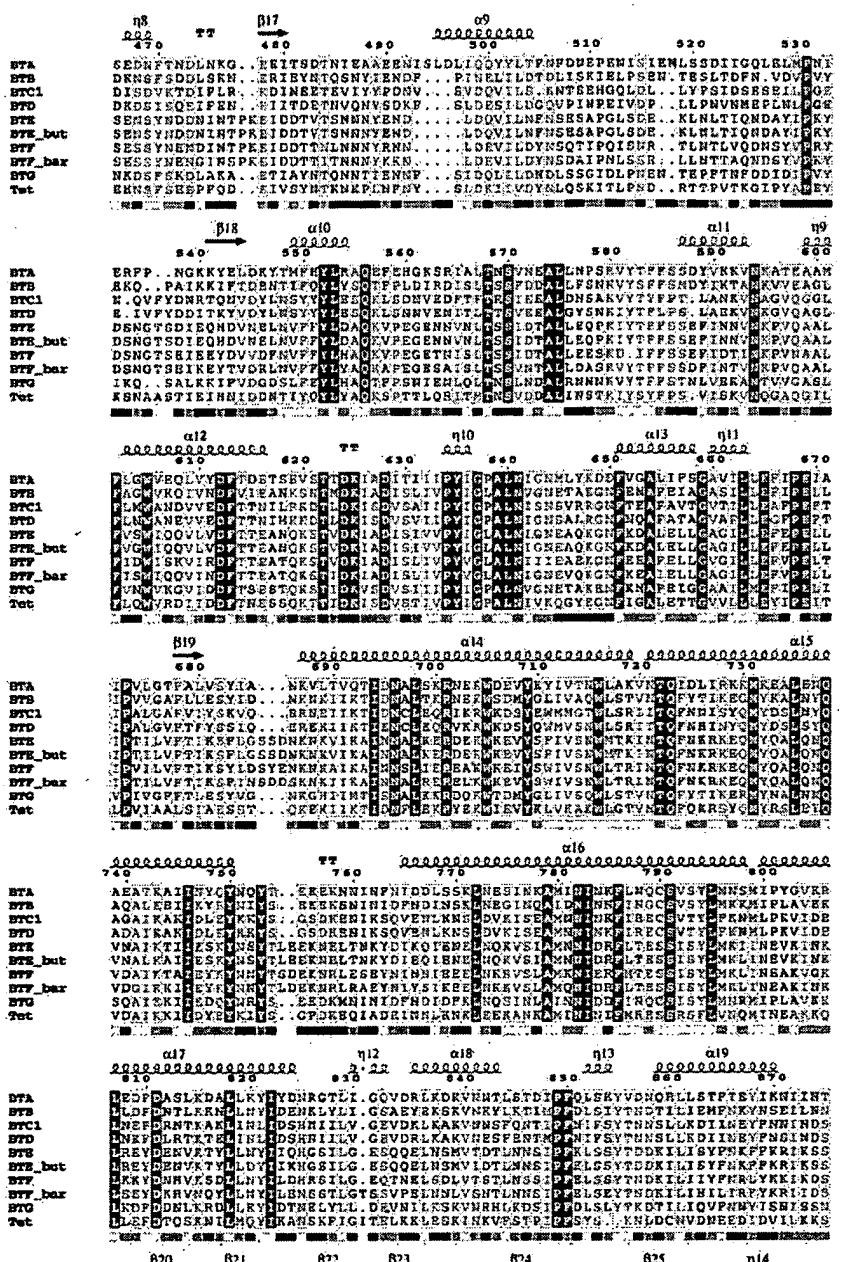

Figure 3 (*legend shown on page 1096*)

inward, seeming to preserve the hydrophobic core of the β-barrel. As most of the antigenicity conferring serotype uniqueness seems to arise in the binding domain (Chen *et al.*, 1997), it is not surprising that its surface residues vary dramatically among the CNTs.

While the existence of a protein receptor binding site is still under investigation, the ganglioside-

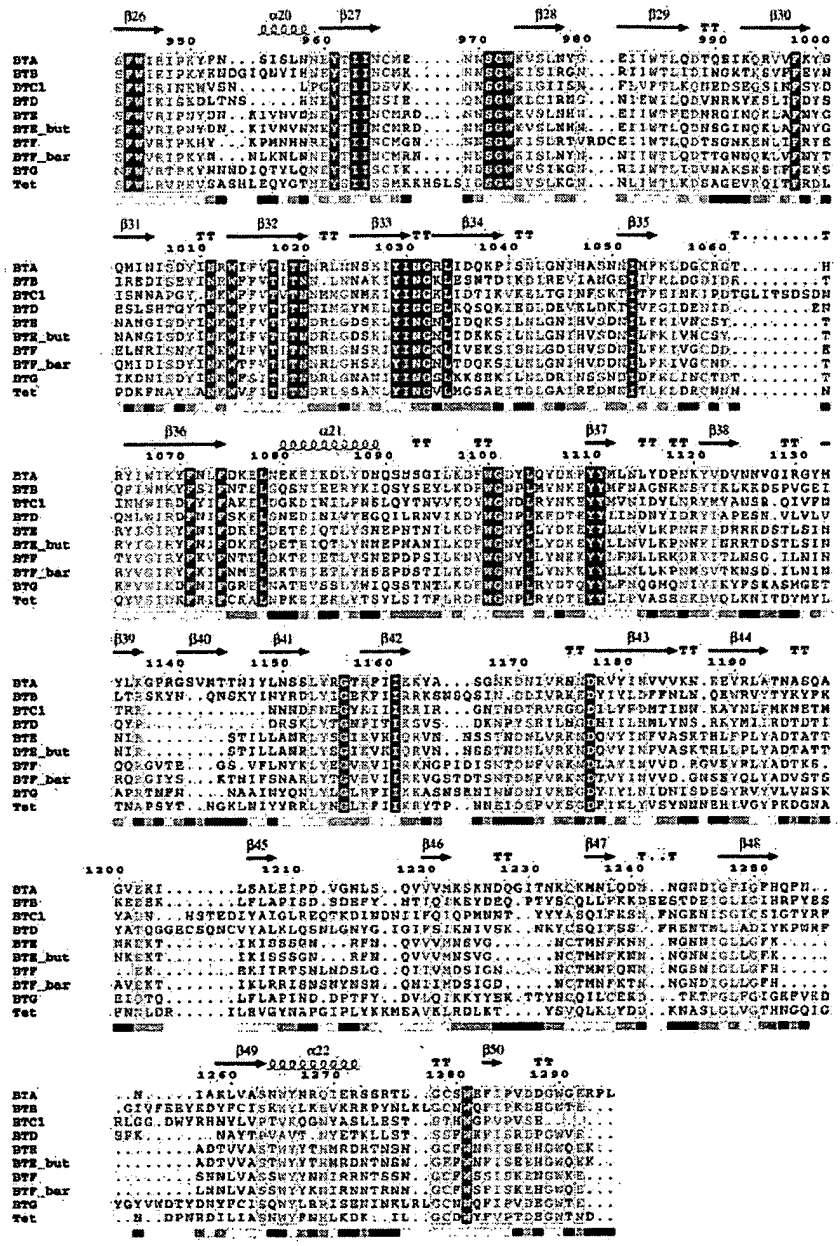

Figure 3 The alignment is displayed to show strict sequence conservation in white letters and a red background, and strong sequence conservation in red letters. The secondary structure elements of the BoNT/A structure are labeled α (α-helix), η (3_{10} helix), β (β-strand) and TT (turn). Assignments were made using DSSP (Kabsch & Sander, 1983) but slightly modified to include helix 20 and strands 31, 40, and 41 after visual inspection. The solvent-accessibility of each residue in the BoNT/A structure is indicated in the bar at the base of the sequences, with white representing buried residues, dark blue representing solvent-accessible residues and light blue representing an intermediate value. This Figure was generated using ESPript (http://www.ipbs.fr/ESPript).

binding site is better understood. A binding study using trypsin digests in BoNT/A indicates that the 30 C-terminal residues of the C-terminal subdomain are involved in binding the GT1b ganglioside (Shone *et al.*, 1985). A photoaffinity labeling study using a novel ganglioside probe implicates this same region in TeNT, showing labeling of H1295 (Shapiro *et al.*, 1997). The functional significance of this region has been identified in BoNT/E (Kubota *et al.*, 1997). A monoclonal antibody capable of neutralizing BoNT/E in mice also bound the peptide YLTHMRD. Comparison of this sequence location in BoNT/E aligns to residues 1292-1298 of TeNT and residues 1266-1272 of

BoNT/A (Figure 3). Most recently, a binding assay that followed tryptophan fluorescence in BoNT/A showed fluorescence quenching upon binding the ganglioside (Kamata *et al.*, 1997). However, with three tryptophan residues within the 30 C-terminal residues it was impossible to narrow the binding site further. The BoNT/A structure shows that of the three tryptophan residues, only one (W1265) is solvent-accessible (Figure 3). This highly hydrophobic residue is fully exposed to solvent and makes contact with Q1269, the residue in BoNT/A that aligns to H1295 of TeNT, implicated in the photoaffinity labeling study. These two residues are located in the i and $i+4$ positions of helix 22, beneath a loop whose orientation varies dramatically between BoNT/A and TeNT (Figures 1(b) and 2(b)). While in BoNT/A this loop points out, creating a deep positively charged cleft between sub-domains, the loop in TeNT folds in, creating a shallow, more neutral cleft. This surface difference could have a role in the alternative localization of the two toxins.

The translocation domain

The translocation domain is able to form channels in artificial bilayers (Blaustein *et al.*, 1987; Donovan & Middlebrook, 1986; Hoch *et al.*, 1985) and in cell membranes (Sheridan, 1998). Visualization of these channels using electron cryomicroscopy suggests that the channels may be formed by the oligomerization of BoNT to form a tetramer (Schmid *et al.*, 1993). Efforts to identify the pore-forming segment(s) of the CNTs have focussed on identifying amphipathic sequences capable of spanning the membrane (Lebeda & Olson, 1995; Montal *et al.*, 1992). Three such sequences were identified using the MOMENT algorithm for hydrophobic moments (595-614, 625-647, and 648-691) (Lebeda & Olson, 1995). A peptide representing part of one of these sequences, 659-681, was shown to form channels in planar lipid bilayers (Oblatt-Montal *et al.*, 1995). This sequence could possibly oligomerize to form a four-helix bundle in the membrane. The structure of BoNT/A, solved at pH 7, does not refute or support this hypothesis. The previously identified amphipathic sequences do not correspond to the long pairs of kinked α-helices observed in the translocation domain. Instead, the sequences precede these helices and adopt primarily extended loop conformations (Figure 4(a)).

The conservation among the CNTs in the translocation domain is shown by sequence in Figure 3 and structurally in Figure 4(c). While virtually no sequence conservation is observed in the region of the belt, the sequence conservation is fairly high throughout the rest of the translocation domain. The majority of the conserved residues are evenly spaced along the long helices (α14, α15, α16, and α17). The organization seems to preserve the packing between each other and between the smaller helices at either end of the domain (α12 and α18). The only area outside of these helices of particular note is the region from 622-644. This

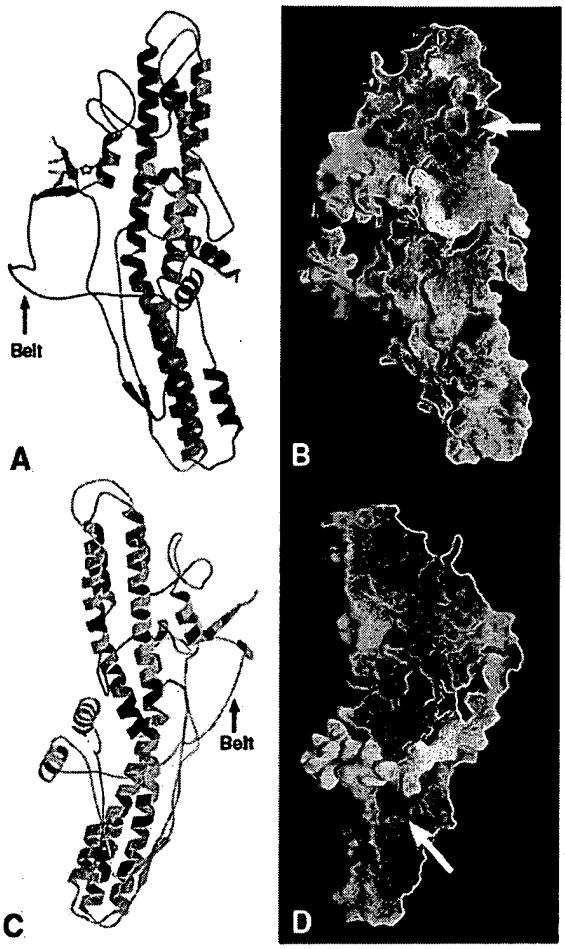

Figure 4. (a) and (b) show identical orientations of the BoNT/A translocation domain and are related to (c) and (d) by a rotation of ~180° around the long axis of the helices. In (c) and (d), the translocation domain belt is protruding out of the plane of the page. In (a), H551 is shown in purple pointing toward the disulfide (yellow) linking the translocation (green) and catalytic (shown partially in blue) domains. The three amphipathic sequences possibly involved in pore formation are colored orange and red, with conserved residues colored red. These sequences form part of a conserved electrostatic cluster shown with an arrow in (b). (c) The overall sequence conservation in the translocation domain structure. The conserved residues, shown in purple, act largely to hold the helical arrangement together, the exception being the conserved residues in the extended loops of the amphipathic sequences, shown in (a). The arrow in (d) points to a cluster of charges conserved at the interface between the main body of the translocation domain and the beginning of the translocation domain belt. (a) and (c) were made using MOLSCRIPT (Kraulis, 1991), while the electrostatic potential surfaces shown in (b) and (d) were made using GRASP (Nicholls *et al.*, 1991).

sequence is largely loop, with one 3_{10} helix (η10) and is part of an amphipathic sequence, 625-647, previously identified (Lebeda & Olson, 1995). The loop extends away from the domain pointing toward the binding domain but is not involved in any inter-domain contacts. To our knowledge, this sequence has not been investigated for its ability to form pores in membranes.

The critical follow-up question after identifying the pore-forming segment is to identify the molecular mechanism by which pH triggers this sequence to change structure and form a membrane-spanning channel. The translocation domain is geared to sense this environmental variable, as the calculated pI for its sequence is 4.66 (as compared to 9.1 for the binding domain and 6.0 for the whole toxin). This low value is attributable to the first half of the domain containing the amphipathic loops (548-685, pI 4.6) as opposed to the second half containing the kinked pair of α-helices (686-872, pI 8.5). The three-dimensional charge distribution indicates several clusters of negative charges on both sides of the surface (Figure 4(b) and (d)). One cluster (Figure 4(b)) is comprised of D612, D615, E616, E619, and E588. Of these residues, D612 is strictly conserved, E616 and E588 are largely conserved, and E619 is consistently polar. The clustering of negative charge is expected to raise the pK_a of these residues, making them titratable at endosomal pH (4.5-5.5). We therefore propose that these residues and possibly any histidine residues will effect the structural changes required for pore formation.

The translocation domain contains only two histidine residues (H551 and H560) located in and after α10, respectively, such that they immediately follow the translocation domain belt (492-545) and precede a disordered loop in the structure. While H560 is solvent-accessible, H551 is buried at the interface of the catalytic and translocation domains within 7 Å of their connecting disulfide. The idea that H551 protonation could disrupt this interface is supported by the fact that the loop following H551 is disordered in the crystal structure. The original data for BoNT/A showed high *B*-factors for residues 560-585, with 561-568 completely disordered. The newer data, collected from a single frozen crystal, revealed density in this area, although the *B*-factors for residues placed in this density remain high (Figure 4(a)). This region likely represents an area of inherent flexibility and is proximal both to the negatively charged cluster and the buried histidine residue. This feature could facilitate the structural changes that accompany pore formation as well as the translocation of the unfolded catalytic domain through the pore. The role that the belt might play in translocation is unknown, although with movement in this loop region it is not implausible to consider that the belt could move as well. A second electrostatically charged cluster on the other side of the translocation domain (Figure 4(d)) might also contribute to belt movement. This cluster is formed by both positive and negative residues of α14 (K700, E703, K704, D706, E707, and K710), along with negative charges at the beginning of the belt (D473, E478, E487, and E490). Protonation at this interface could interrupt interactions holding the belt to the main body of the translocation domain thus allowing for the translocation and release of the catalytic domain. While the domains are tethered by a disulfide and the interface between the catalytic and translocation domains (not including the belt) includes ~1190 Å^2 of buried surface area, the non-covalent contacts between the belt and catalytic domain are not substantial. The 54 residue belt buries an additional ~1620 Å^2 of the catalytic domain surface area. This translates to ~30 Å^2 per residue of the belt, considerably lower than the ~200 Å^2 buried by an isoleucine residue in a well-packed protein core. The energy required for removing the belt is probably even lower than that indicated by the buried surface area, since the belt is presumably unstructured in the absence of the catalytic domain, thus adding an entropic cost to their association.

The catalytic domain

Previous alignments have indicated that the catalytic domains share up to 36.5% sequence identity, the exception being the catalytic domains of BoNT/B and TeNT, which share 51.6% sequence identity (Kurazono *et al.*, 1992). All sequences contain a HEXXH sequence motif, typical of many zinc proteases. The catalytic domain sequence alignment also shows the strict conservation of two Asp-Pro bonds, one immediately preceding the HEXXH sequence (D215 and P216) and the other at the N terminus of the molecule (D11 and P12). A third Asp-Pro bond is present in all of the sequences, except BoNT/A, where S74 exists instead of proline. Asp-Pro bonds from BoNT/A have been shown to hydrolyze at pH 5 (DasGupta & Evenson, 1992; DasGupta & Tepp, 1991), a fact that may bear relevance given the low pH environment of the endosome and the need to translocate a 50 kDa protein. It should be noted, however, that a truncation after residue 9 inactivates *in vitro* catalysis (Kurazono *et al.*, 1992).

Other than the zinc protease HEXXH sequence motif, the catalytic domains share no sequence similarity with proteins outside of the CNT family. While a DALI search showed that thermolysin, a well-characterized zinc protease, bore the highest structural identity with the BoNT/A catalytic domain, this similarity was weak with a Z-score of 4.6 for 139 residues (Lacy *et al.*, 1998). Nevertheless, a visual comparison of the two enzymes (Figure 5) is useful, because, like BoNT/A, thermolysin coordinates the active-site zinc ion with two

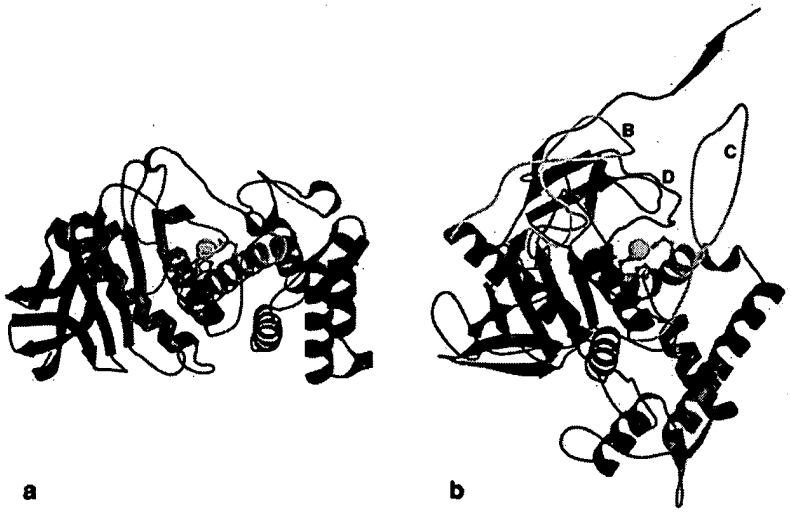

Figure 5. A comparison of (a) thermolysin and (b) the BoNT/A catalytic domain. The helix containing the HEXXH zinc protease motif is colored red, while the helix presenting the second Glu is colored gold. The topology of the secondary structural elements differs such that in BoNT/A, the shorter green and gold helices come together to form the structural equivalent of the single gold helix in thermolysin. Loops B, C, and D of BoNT/A are colored yellow, pink, and orchid, respectively, and form boundaries of accessibility to the active-site zinc ion.

histidine residues (His142 and His146), a glutamate residue (Glu166), and a water-mediated glutamate residue (Glu143) (Colman *et al.*, 1972). Structural similarities include the helix containing the HEXXH sequence (α3) and a four-stranded β-sheet (β3, β6, β7, and β8). This accounts for the structural presentation of the two histidine ligands and the glutamate residue that coordinates the activated water in catalysis. The presentation of the fourth ligand is intriguing, as thermolysin presents the Glu166 ligand with a single helix, while BoNT/A has generated a similar presentation by folding with two smaller helices (α4 and α8) end-to-end. As the two BoNT/A helices point in opposite directions, it is uncertain if maintaining this helical periodicity plays a role in the Glu261 presentation. More interestingly, it supports this activity as a likely example of convergent evolution.

The fold comparison highlights the likely cleft by which the substrate gains accessibility to the active-site. This channel of accessibility is bordered by three loops in BoNT/A designated B (47-80), C (231-259), and D (356-371) (Figure 5(b)). The loops in BoNT/A are longer than those in thermolysin and result in a more buried active-site, although these loops could be flexible or oriented differently in the absence of the translocation domain. A second avenue of active-site access would be for substrate to enter from a channel seen arially both in Figure 5(b) and Figure 6(a). The presence of the translocation domain belt and the main body of the translocation domain (Figure 6(b)) occlude both channels of accessibility in the holotoxin.

The conserved residues of the catalytic domain were mapped onto the three-dimensional structure of BoNT/A and are shown in Figure 7. Conserved residues within 6 Å of the active-site cavity are depicted with their full side-chains. In addition to the anticipated residues of the HEXXH motif (H222, E223, and H226), both E260 and E261

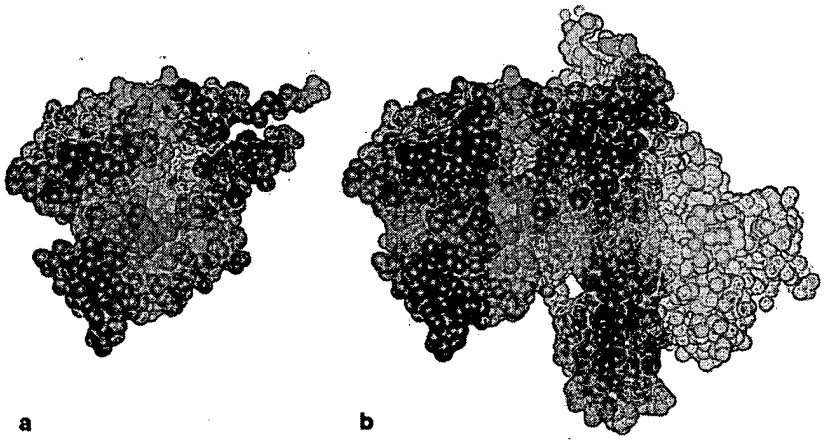

Figure 6. A CPK model of (a) the BoNT/A catalytic domain and (b) the BoNT/A holotoxin showing how accessibility to the buried active-site zinc ion (orange) is occluded both by the translocation domain belt and the long axis of the translocation domain (green).

Figure 7. The BoNT/A catalytic domain colored to show residues with high sequence conservation in purple. Seven strictly conserved residues within 6 Å of the active-site zinc ion (gray sphere) are colored orange and include H222, E223, H226, E260, E261, R362, and Y365.

(where E261 can coordinate the zinc ion) and R362 and Y365 of loop D are located within this proximity. These seven residues are located in a plane at the base of the active-site channel, thus making it reasonable to assume that their conservation preserves the general zinc endopeptidase activity. The specificity is therefore likely to arise from the residues forming the channel, or cavity, above this base. This cavity is at least 1075 $Å^3$ in the presence of the translocation domain. Non-conserved catalytic domain residues within 4 Å of this cavity include 65-68 of loop B, 161-163, 193, 219, 238 of loop C, 255-258, and 368-9 of loop D. These residues were analyzed in smaller subsets, particularly BoNT/B and TeNT, for sequence conservation, but none was readily apparent. Residues 459, 530, 531, 533, and 536 of the translocation domain also border the active-site cavity where P531 is conserved in all ten sequences and is the only strictly conserved residue of the belt. These residues, which occlude access to the active-site in the holotoxin form, are presumably absent following translocation through the membrane. The catalytic domain may adopt an altered conformation following its translocation, a possibility that should be addressed by the structure determination of this domain in the absence of the neurotoxin.

While some level of sequence conservation was expected in the vicinity of the active-site, the appearance of conservation in two surface loops, distant from the active-site was surprising. Loop A (1-18) is comprised of conserved residues F7, N8, Y9, D11, and P12, and is close to conserved residues of β1, β2, β3, β6, as well as the N terminus of α1 (Figure 8). The importance of this loop was identified in experiments measuring the toxicity of N-terminal and C-terminal catalytic domain deletions (Kurazono *et al.*, 1992). While a deletion of residues 1-7 was tolerated, the deletion of residues 1-9 was not, implicating a direct biochemical and/or indirect structural role for one or both of residues N8 and Y9. While N8 is largely solvent-accessible, Y9 forms part of a conserved charged pocket

Figure 8. The BoNT/A catalytic domain is shown in purple with residues 1-88 colored yellow. Residues 450-550 of the translocation domain are colored green to show the relative position of the translocation domain belt. The Asp-Pro residues are colored red, while the conserved cluster of residues around Tyr9 is shown in purple. The backbone corresponding to the three proline residues P60-P62 is colored green and shows the relative proximity to the active-site zinc ion (gray sphere). Labels are underneath their corresponding secondary structural element.

containing K33 (β2), E46 (β3), D80 (α1), and K83 (α1) (Figure 8). Since D80 and K83 are contained within one end of α1 and the other end of α1 has contact with α3, the HEXXH helix, it is possible that this conservation is merely to preserve the orientation of these helices, and therefore, the structural integrity of the catalytic machinery. This possibility seems unlikely, however, given the 30 Å distance of Y9 from the zinc ion and the lack of sequence conservation in secondary structural elements more proximal to the HEXXH helix (for example, β7 and β8). It seems more likely that the conserved presence of Tyr in the charged pocket is preserving a remote tertiary structure critical for catalysis. This is supported by experiments showing that a monoclonal antibody (mAb) mapping to residues 27-52, when administered intraneurally, neutralized the effects of BoNT/A (Cenci Di Bello *et al.*, 1994). The solvent-accessible residues of this sequence and likely mAb binding site immediately preceed and follow the strands β2 and β3, respectively.

The second conserved surface loop, the last part of loop B, is also located within this proximity. It is intriguing that both loops A and B contain an Asp-Pro pair of residues, although the conservation in loop B did not extend to BoNT/A (Figure 8). While acidic pH is not required for catalysis, cleavage in these sites may aid the efficiency of translocation. The conservation in the latter part of loop B is also of interest, as the preceding residues come into contact with the active-site cavity. An unusual sequence of three consecutive proline residues (P60, P61, and P62) is followed by residues 65-68 (previously noted for their proximity to the channel).

There is mounting evidence suggesting the existence of a remote recognition sequence necessary for catalysis. A ten residue sequence shared by all three of the SNARE proteins has been identified and termed the SNARE motif (Rossetto *et al.*, 1994). This motif is thought to be required for efficient catalysis and is present at varying distances upstream of the cleavage site. A potentially related study shows that the addition of peptides *in trans* upstream and downstream of the TeNT cleavage site is able to activate the toxin (Cornille *et al.*, 1997). The idea of exosite-dependent cleavage implies that a con-

formational change, in addition to simple removal of the translocation domain, may be required for catalysis. Perhaps the regions in BoNT/A that share high degrees of sequence conservation distal from the active-site represent such general activation domains. Alternatively, these conserved surface loops could represent oligomeric interfaces or be involved in localization of the toxin within the cytosol.

Materials and Methods

Sequence alignments and superposition on the BoNT/A structure

Nine CNT sequences were aligned to that of BoNT/A using CLUSTAL W (Thompson *et al.*, 1994). A gap opening penalty of 10, gap extension penalty of 0.05, and gap separation distance of 8 were used with the BLOSUM62 matrix (Henikoff & Henikoff, 1996). This matrix was also used in the display in Figure 3, using a global score of 0.15. The Risler matrix was used to generate per residue scores for the alignment (Risler *et al.*, 1988). These scores were output in the *B*-factor column of the BoNT/A coordinates, facilitating the display of sequence conservation on the three-dimensional structure. Residues with scores 68 and higher were colored purple, while those below 68 were colored yellow. The number 68 was empirically determined to best match the display of sequence conservation seen in Figure 3.

Toxin structure at −170°C

The protein was concentrated to 5 mg/ml in 10 mM Hepes (pH 7), 100 mM KCl. This stock was mixed 1:1 (v/v) with a mother liquor of 30% (v/v) PEG 600 and 100 mM Tris (pH 7) and allowed to crystallize by hanging drop vapor diffusion. The crystal formed overnight and grew to a size of about 100 μm × 200 μm × 520 μm over the course of a week. The crystal was washed briefly in a solution made from 2 μl of mother liquor and 2 μl of 30% PEG 600 and plunged into liquid nitrogen. Data were collected at SSRL beamline 9-1 and processed with DENZO, yielding data 99.4% complete to 3.3 Å with 98.5% completeness in the last bin. The data were 5.8-fold redundant with an R_{merge} of 8.5% overall. After orienting the molecule with AMoRe, the structure was readily refined into the frozen data using CNS, yielding a final R_{work} of 26% and R_{free} of 31%. While the processing and refinement statistics indicate that the quality is lower than the structure solved at 4°C, these data did bring in density for residues 561-568.

References

Aureli, P., Fenicia, L., Pasolini, B., Gianfranceschi, M., McCroskey, L. M. & Hatheway, C. L. (1986). Two cases of type E infant botulism caused by neurotoxigenic *Clostridium butyricum* in Italy. *J. Infect. Dis.* **154**, 207-211.

Binz, T., Blasi, J., Yamasaki, S., Baumeister, A., Link, E., Sudhof, T. C., Jahn, R. & Niemann, H. (1994). Proteolysis of SNAP-25 by types E and A botulinal neurotoxins. *J. Biol. Chem.* **269**, 1617-1620.

Black, J. D. & Dolly, J. O. (1986a). Interaction of ^{125}I-labeled botulinum neurotoxins with nerve terminals. I. Ultrastructural autoradiographic localization and quantitation of distinct membrane acceptors for types A and B on motor nerves. *J. Cell. Biol.* **103**, 521-534.

Black, J. D. & Dolly, J. O. (1986b). Interaction of ^{125}I-labeled botulinum neurotoxins with nerve terminals. II. Autoradiographic evidence for its uptake into motor nerves by acceptor-mediated endocytosis. *J. Cell. Biol.* **103**, 535-544.

Blasi, J., Chapman, E. R., Link, E., Binz, T., Yamasaki, S., De Camilli, P., Sudhof, T. C., Niemann, H. & Jahn, R. (1993a). Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25. *Nature*, **365**, 160-163.

Blasi, J., Chapman, E. R., Yamasaki, S., Binz, T., Niemann, H. & Jahn, R. (1993b). Botulinum neurotoxin C1 blocks neurotransmitter release by means of cleaving HPC-1/syntaxin. *EMBO J.* **12**, 4821-4828.

Blaustein, R. O., Germann, W. J., Finkelstein, A. & DasGupta, B. R. (1987). The N-terminal half of the heavy chain of botulinum type A neurotoxin forms channels in planar phospholipid bilayers. *FEBS Letters*, **226**, 115-120.

Cenci Di Bello, I., Poulain, B., Shone, C. C., Tauc, L. & Dolly, J. O. (1994). Antagonism of the intracellular action of botulinum neurotoxin type A with monoclonal antibodies that map to light-chain epitopes. *Eur. J. Biochem.* **219**, 161-169.

Chen, F., Kuziemko, G. M., Amersdorfer, P., Wong, C., Marks, J. D. & Stevens, R. C. (1997). Antibody mapping to domains of botulinum neurotoxin serotype A in the complexed and uncomplexed forms. *Infect. Immun.* **65**, 1626-1630.

Colman, P. M., Jansonius, J. N. & Matthews, B. W. (1972). The structure of thermolysin: an electron density map at 2-3 Å resolution. *J. Mol. Biol.* **70**, 701-724.

Cornille, F., Martin, L., Lenoir, C., Cussac, D., Roques, B. P. & Fournie-Zaluski, M. C. (1997). Cooperative exosite-dependent cleavage of synaptobrevin by tetanus toxin light chain. *J. Biol. Chem.* **272**, 3459-3464.

DasGupta, B. R. & Evenson, M. L. (1992). Botulinum neurotoxin type A: the isolated light chain breaks down. *FASEB J.* **6**, A227.

DasGupta, B. R. & Tepp, W. (1991). Botulinum neurotoxin type A breaks down at pH 5.0 at the two Asp-Pro bonds of the heavy chain. *Soc. Neurosci. Abstr.* **17**, 1526.

Dolly, J. O., Black, J., Williams, R. S. & Melling, J. (1984). Acceptors for botulinum neurotoxin reside on motor nerve terminals and mediate its internalization. *Nature*, **307**, 457-460.

Donovan, J. J. & Middlebrook, J. L. (1986). Ion-conducting channels produced by botulinum toxin in planar lipid membranes. *Biochemistry*, **25**, 2872-2876.

Foran, P., Lawrence, G. W., Shone, C. C., Foster, K. A. & Dolly, J. O. (1996). Botulinum neurotoxin C1 cleaves both syntaxin and SNAP-25 in intact and permeabilized chromaffin cells: correlation with its blockade of catecholamine release. *Biochemistry*, **35**, 2630-2636.

Hall, J. D., McCroskey, L. M., Pincomb, B. J. & Hatheway, C. L. (1985). Isolation of an organism resembling *Clostridium barati* which produces type F

botulinal toxin from an infant with botulism. *J. Clin. Microbiol.* **21**, 654-655.

Henikoff, J. G. & Henikoff, S. (1996). Blocks database and its applications. *Methods Enzymol.* **266**, 88-105.

Hoch, D. H., Romero-Mira, M., Ehrlich, B. E., Finkelstein, A., DasGupta, B. R. & Simpson, L. L. (1985). Channels formed by botulinum, tetanus, and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes. *Proc. Natl Acad. Sci. USA*, **82**, 1692-1696.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kamata, Y., Yoshimoto, M. & Kozaki, S. (1997). Interaction between botulinum neurotoxin type A and ganglioside: ganglioside inactivates the neurotoxin and quenches its tryptophan fluorescence. *Toxicon*, **35**, 1337-1340.

Kraulis, P. J. (1991). MOLSCRIPT- a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kubota, T., Watanabe, T., Yokosawa, N., Tsuzuki, K., Indoh, T., Moriishi, K., Sanda, K., Maki, Y., Inoue, K. & Fujii, N. (1997). Epitope regions in the heavy chain of *Clostridium botulinum* type E neurotoxin recognized by monoclonal antibodies. *Appl. Environ. Microbiol.* **63**, 1214-1218.

Kurazono, H., Mochida, S., Binz, T., Eisel, U., Quanz, M., Grebenstein, O., Wernars, K., Poulain, B., Tauc, L. & Niemann, H. (1992). Minimal essential domains specifying toxicity of the light chains of tetanus toxin and botulinum neurotoxin type A. *J. Biol. Chem.* **267**, 14721-14729.

Lacy, D. B. & Stevens, R. C. (1998). Unraveling the structures and modes of action of bacterial toxins. *Curr. Opin. Struct. Biol.* **8**, 778-784.

Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. & Stevens, R. C. (1998). Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nature Struct. Biol.* **5**, 898-902.

Lebeda, F. J. & Olson, M. A. (1994). Secondary structural predictions for the clostridial neurotoxins. *Proteins: Struct. Funct. Genet.* **20**, 293-300.

Lebeda, F. J. & Olson, M. A. (1995). Structural predictions of the channel-forming region of botulinum neurotoxin heavy chain. *Toxicon*, **33**, 559-567.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Montal, M. S., Blewitt, R., Tomich, J. M. & Montal, M. (1992). Identification of an ion channel-forming motif in the primary structure of tetanus and botulinum neurotoxins. *FEBS Letters*, **313**, 12-18.

Montecucco, C. (1986). How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem. Sci.* **11**, 315-317.

Montecucco, C. & Schiavo, G. (1995). Structure and function of tetanus and botulinum neurotoxins. *Quart. Rev. Biophys.* **28**, 423-472.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

Oblatt-Montal, M., Yamazaki, M., Nelson, R. & Montal, M. (1995). Formation of ion channels in lipid bilayers by a peptide with the predicted transmembrane sequence of botulinum neurotoxin A. *Protein Sci.* **4**, 1490-1497.

Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988). Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* **204**, 1019-1029.

Rossetto, O., Schiavo, G., Montecucco, C., Poulain, B., Deloye, F., Lozzi, L. & Shone, C. C. (1994). SNARE motif and neurotoxins. *Nature*, **372**, 415-416.

Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B. R. & Montecucco, C. (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature*, **359**, 832-835.

Schiavo, G., Rossetto, O., Catsicas, S., Polverino de Laureto, P., DasGupta, B. R., Benfenati, F. & Montecucco, C. (1993a). Identification of the nerve terminal targets of botulinum neurotoxin serotypes A, D, and E. *J. Biol. Chem.* **268**, 23784-23787.

Schiavo, G., Santucci, A., Dasgupta, B. R., Mehta, P. P., Jontes, J., Benfenati, F., Wilson, M. C. & Montecucco, C. (1993b). Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds. *FEBS Letters*, **335**, 99-103.

Schiavo, G., Shone, C. C., Rossetto, O., Alexander, F. C. & Montecucco, C. (1993c). Botulinum neurotoxin serotype F is a zinc endopeptidase specific for VAMP/synaptobrevin. *J. Biol. Chem.* **268**, 11516-11519.

Schiavo, G., Malizio, C., Trimble, W. S., Polverino de Laureto, P., Milan, G., Sugiyama, H., Johnson, E. A. & Montecucco, C. (1994). Botulinum G neurotoxin cleaves VAMP/synaptobrevin at a single Ala-Ala peptide bond. *J. Biol. Chem.* **269**, 20213-20216.

Schmid, M. F., Robinson, J. P. & DasGupta, B. R. (1993). Direct visualization of botulinum neurotoxin-induced channels in phospholipid vesicles. *Nature*, **364**, 827-830.

Shapiro, R. E., Specht, C. D., Collins, B. E., Woods, A. S., Cotter, R. J. & Schnaar, R. L. (1997). Identification of a ganglioside recognition domain of tetanus toxin using a novel ganglioside photoaffinity ligand. *J. Biol. Chem.* **272**, 30380-30386.

Sheridan, R. E. (1998). Gating and permeability of ion channels produced by botulinum toxin types A and E in PC12 cell membranes. *Toxicon*, **36**, 703-717.

Shone, C. C., Hambleton, P. & Melling, J. (1985). Inactivation of *Clostridium botulinum* type A neurotoxin by trypsin and purification of two tryptic fragments. Proteolytic action near the COOH- terminus of the heavy subunit destroys toxin-binding activity. *Eur. J. Biochem.* **151**, 75-82.

Simpson, L. L. (1980). Kinetic studies on the interaction between botulinum toxin type A and the cholinergic neuromuscular junction. *J. Pharmacol. Expt. Ther.* **212**, 16-21.

Simpson, L. L. (1989). *Botulinum Neurotoxin and Tetanus Toxin*, Academic Press, San Diego.

Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature*, **395**, 347-353.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.

Umland, T. C., Wingert, L. M., Swaminathan, S., Furey, W. F., Schmidt, J. J. & Sax, M. (1997). Structure of

the receptor binding fragment HC of tetanus neurotoxin. *Nature Struct. Biol.* **4**, 788-792.

Yamasaki, S., Binz, T., Hayashi, T., Szabo, E., Yamasaki, N., Eklund, M., Jahn, R. & Niemann, H. (1994a). Botulinum neurotoxin type G proteolyses the Ala81-Ala82 bond of rat synaptobrevin 2. *Biochem. Biophys. Res. Commun.* **200**, 829-835.

Yamasaki, S., Hu, Y., Binz, T., Kalkuhl, A., Kurazono, H., Tamura, T., Jahn, R., Kandel, E. & Niemann, H. (1994b). Synaptobrevin/vesicle-associated membrane protein (VAMP) of *Aplysia californica*: structure and proteolysis by tetanus toxin and botulinal neurotoxins type D and F. *Proc. Natl Acad. Sci. USA*, **91**, 4688-4692.

Edited by G. Von Heijne

(Received 29 March 1999; received in revised form 8 June 1999; accepted 9 June 1999)